

CONVERGENCE OF DISCRETE TIME KALMAN FILTER ESTIMATE TO CONTINUOUS TIME ESTIMATE*

ATTE AALTO^{a,b}

^aDepartment of Mathematics and Systems Analysis, Aalto University, Espoo, Finland

^bInria, Université Paris–Saclay, Palaiseau, France; MÆDISIM team

ABSTRACT. This article is concerned with the convergence of the state estimate obtained from the discrete time Kalman filter to the continuous time estimate as the temporal discretization is refined. The convergence follows from Martingale convergence theorem as demonstrated below but, surprisingly, no results exist on the rate of convergence. We derive convergence rate estimates for the discrete time Kalman filter estimate for finite and infinite dimensional systems. The proofs are based on applying the discrete time Kalman filter on a dense numerable subset of a certain time interval $[0, T]$.

Keywords: Kalman filter, infinite dimensional systems, temporal discretization, sampled data

1. INTRODUCTION

It is well known that Kalman filter (or Kalman–Bucy filter) gives the optimal solution to the state estimation problem for discrete (or continuous) time linear systems with Gaussian initial state, and Gaussian input and output noise processes. These filters have proven to be robust and they have been widely used in practical applications since their introduction in the 1960s. The implementation of the discrete time filter is straightforward since it is readily formulated in an algorithmic manner. Thus, it may often be tempting to use the discrete time filter on the temporally discretized continuous time system. The purpose of this article is to study the convergence of a state estimate from discrete time Kalman filter to the continuous time state estimate as the temporal discretization is refined. In particular, we show convergence speed estimates for the quadratic error between the discrete time and continuous time estimate first for finite dimensional systems without input noise, then finite dimensional systems with input noise, and finally, for infinite dimensional systems with a bounded observation operator.

The class of systems studied here is described by mappings (A, B, C) where $A : \mathcal{X} \rightarrow \mathcal{X}$, $B : \mathcal{U} \rightarrow \mathcal{X}$, and $C : \mathcal{X} \rightarrow \mathcal{Y}$, and the corresponding

*This is the accepted author's version of the manuscript accepted for publication in International Journal of Control.

dynamics equations

$$(1) \quad \begin{cases} dz(t) = Az(t) dt + Bdu(t), & t \in \mathbb{R}^+, \\ dy(t) = Cz(t) dt + dw(t), \\ z(0) = x. \end{cases}$$

Here \mathcal{X} is called the *state space*, $\mathcal{U} = \mathbb{R}^q$ is the *input space*, and $\mathcal{Y} = \mathbb{R}^r$ is the *output space*. The mapping A is the generator of a contractive C_0 -semigroup e^{At} on \mathcal{X} with domain $\mathcal{D}(A)$, $B : \mathbb{R}^q \rightarrow \mathcal{X}$ is the *input operator*, and $C : \mathcal{X} \rightarrow \mathbb{R}^r$ is called the *observation operator*. The observation operator can be bounded or not but it always maps to a finite dimensional space in this article. The process y is called the *output process*. The *input* and *output noise processes* u and w are assumed to be q - and r -dimensional Brownian motions with incremental covariance matrices $Q > 0$ and $R > 0$, respectively, and the *initial state* x is assumed to be an \mathcal{X} -valued Gaussian random variable, $x \sim N(m, P_0)$. The noise processes u and w and the initial state x are assumed to be mutually independent. Note that the system (1) is written as a stochastic differential equation. For background of stochastic equations and the formulation of the Kalman–Bucy filter in this framework, we refer to [13] (in particular, Section 6.3 therein) and [6].

The discrete and continuous time state estimates are defined by

$$(2) \quad \hat{z}_{T,n} := \mathbb{E}\left(z(T) \mid \left\{y\left(\frac{iT}{n}\right)\right\}_{i=1}^n\right) \quad \text{and} \quad \hat{z}(T) := \mathbb{E}\left(z(T) \mid \{y(s), s \leq T\}\right),$$

respectively. That is, we are estimating the final state of the system (1). These estimates are given by the discrete and continuous time Kalman filter, respectively. The purpose of this article is to study the convergence $\hat{z}_{T,n} \rightarrow \hat{z}(T)$ as $n \rightarrow \infty$.

In Section 2, we cover the necessary background concerning stochastics and the Kalman filter. The proofs of the main results are based on using the discrete time Kalman filter on a sequence that forms a dense subset of the interval $[0, T]$. In particular, in Section 2.1, it is shown that this procedure in fact converges to $\hat{z}(T)$ strongly in \mathcal{X} almost surely. Gaussian random variables and the Kalman filter are introduced in Section 2.2. Section 3 contains the main result in the simplest case, namely an estimate of the convergence speed of $\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right)$ when n is increased for finite dimensional system without input noise. The proofs of the other results follow the same outline, and so this simplest case is shown in full detail in order to convey the ideas as clearly as possible. In the beginning of the section it is shown how to take into account an intermediate measurement in Kalman filtering — an important tool in the proofs. The result for systems with input noise is shown in Section 4 and for infinite dimensional systems with bounded observation operator the result is generalized in Section 5.

The Kalman filter performance has been widely studied in literature. Even though it was originally derived for state estimation for finite dimensional linear systems with Gaussian input and output noise processes it has proven to be very robust and thus applicable to a variety of other scenarios. Variants for non-linear systems have been developed, such as the extended Kalman filter and the unscented Kalman filter, see the book [16]. Kalman

filter sensitivity to modelling errors has been studied by for example [17] and [7, Chapter 7]. See also the recent work [12] for a study on the effect of modelling errors in an infinite dimensional example case, namely the one dimensional wave equation. The effect of state space discretization to Kalman filtering has been studied in, *e.g.*, [3], [8], and in [1].

However, the error that stems from using the discrete time filter on the temporally discretized continuous time system has not received much attention. Two recent articles, [2] and [18], have studied different numerical methods for approximating the matrix exponential $e^{A\Delta t}$ and the effect of this approximation on the solution of the corresponding Lyapunov equations and Kalman filtering. A convergence result of the discrete time Kalman filter estimate in finite dimensional setting is shown by [15] without convergence rate estimate. They use similar techniques that can also be used to (formally) obtain the Kalman-Bucy filter as a limit of the discrete time Kalman filter, as is done for example in [16, Section 8.2] and [7, Section 4.3].

Notation and standing assumptions.

- The space of bounded operators from a Hilbert space \mathcal{H}_1 to another Hilbert space \mathcal{H}_2 is denoted by $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, and $\mathcal{L}(\mathcal{H}_1) = \mathcal{L}(\mathcal{H}_1, \mathcal{H}_1)$.
- We assume that the state space \mathcal{X} is a separable Hilbert space. Denote by $\{e_k\}_{k=1}^{p/\infty} \subset \mathcal{X}$ an orthonormal basis for the p/∞ -dimensional state space.
- A is the generator of a C_0 -semigroup on \mathcal{X} . The semigroup is denoted by e^{At} even though A is not bounded in general. We assume $\|e^{At}\|_{\mathcal{L}(\mathcal{X})} \leq \mu$ for $t \in [0, T]$.
- The space $\mathcal{D}(A)$ is equipped with the graph norm $\|x\|_{\mathcal{D}(A)}^2 = \|x\|_{\mathcal{X}}^2 + \|Ax\|_{\mathcal{X}}^2$ which makes $\mathcal{D}(A)$ a Hilbert space since A is closed.
- We assume that the observation operator is bounded, $C \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, and that the input operator is smooth, that is, $B \in \mathcal{L}(\mathcal{U}, \mathcal{D}(A))$. The input and output spaces are always finite dimensional, $\mathcal{U} = \mathbb{R}^q$ and $\mathcal{Y} = \mathbb{R}^r$.
- Ω is a probability space and $L^2(\Omega; \mathcal{X})$ is the space of \mathcal{X} -valued random variables ξ satisfying $\mathbb{E}(\|\xi\|_{\mathcal{X}}^2) < \infty$.
- The sigma algebra generated by a random variable ξ is denoted by $\sigma\{\xi\}$.
- To improve readability, we use index n only when referring to the discretization level in the state estimate $\hat{z}_{T,n}$ defined in (2), index k only to denote different dimensions of the state space, and index j only when referring to the martingale \tilde{z}_j defined below in Section 2.1.

2. BACKGROUND AND PRELIMINARY RESULTS

As mentioned above, the proofs of this article are based on applying the discrete time Kalman filter on a dense, numerable subset on the interval $[0, T]$ — starting from the discrete time state estimate $\hat{z}_{T,n}$ — and computing an upper bound for the change in the estimate. In section 2.1, we establish that the limit thus obtained is indeed $\hat{z}(T)$. Gaussian random variables and the Kalman filter are discussed in Section 2.2.

2.1. Stochastics. In the cases where the state space \mathcal{X} is infinite dimensional it is always assumed either that $x \in \mathcal{D}(A)$ almost surely or that $C \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. This guarantees that the stochastic process y given by (1) has almost surely continuous sample paths. Let $\{t_i\}_{i=1}^\infty$ be a dense subset of the interval $[0, T]$ and denote $T_j := \{t_i\}_{i=1}^j$. Now let ξ be an integrable \mathcal{X} -valued random variable and y a stochastic process with almost surely continuous sample paths. Then $[\xi]_k := \langle \xi, e_k \rangle_{\mathcal{X}}$ is an integrable \mathbb{R} -valued random variable for each k . Define the martingales $[\tilde{\xi}_j]_k := \mathbb{E}(\langle \xi, e_k \rangle_{\mathcal{X}} | \mathcal{F}_j)$ where \mathcal{F}_j is the sigma algebra generated by $\{y(t), t \in T_j\}$, that is, $\mathcal{F}_j = \sigma\{y(t), t \in T_j\}$. It holds that $\mathbb{E}(|[\tilde{\xi}_j]_k|) \leq \mathbb{E}(|\langle \xi, e_k \rangle_{\mathcal{X}}|)$ for all j and thus by Doob's Martingale convergence theorem (see [13, Appendix C], in particular, Theorem C.6 and Corollary C.9), $[\tilde{\xi}_j]_k \rightarrow [\tilde{\xi}_\infty]_k$ almost surely. As y has continuous sample paths, it holds that $[\xi_\infty]_k = \mathbb{E}(\langle \xi, e_k \rangle_{\mathcal{X}} | \{y(s), s \leq T\})$ almost surely. Using this componentwise implies that $\tilde{\xi}_j := \mathbb{E}(\xi | \mathcal{F}_j) = \sum_{k=1}^\infty [\tilde{\xi}_j]_k e_k$ converges strongly (in \mathcal{X}) almost surely to $\tilde{\xi}_\infty = \sum_{k=1}^\infty [\tilde{\xi}_\infty]_k e_k$.

In general, the martingale convergence theorem is true for Banach spaces that have the Radon–Nikodym property. All reflexive Banach spaces (and therefore also Hilbert spaces) have the Radon–Nikodym property. The above deduction follows essentially the proof of this fact in the special case of \mathcal{X} being a Hilbert space, see [14, Corollary 2.11].

In the proofs, we will need the following telescope identity for martingales.

Lemma 1. *Let ξ_j be a square integrable \mathcal{X} -valued martingale. Then for $L, N \in \mathbb{N}$ with $L \geq N$:*

$$\mathbb{E}(\|\xi_L - \xi_N\|_{\mathcal{X}}^2) = \sum_{j=N}^{L-1} \mathbb{E}(\|\xi_{j+1} - \xi_j\|_{\mathcal{X}}^2).$$

Proof. The result follows directly from the fact that martingale increments are orthogonal. Let us show this. Let $k \geq j$ and denote $\mathcal{F}_i = \sigma\{\xi_1, \dots, \xi_i\}$. Then (recalling that $\xi_k - \mathbb{E}(\xi_k | \mathcal{F}_j) \perp \xi_i$ for $i \leq j$ and the martingale property $\mathbb{E}(\xi_k | \mathcal{F}_j) = \xi_j$),

$$\mathbb{E}(\langle \xi_k, \xi_j \rangle_{\mathcal{X}}) = \mathbb{E}(\langle \mathbb{E}(\xi_k | \mathcal{F}_j) + (\xi_k - \mathbb{E}(\xi_k | \mathcal{F}_j)), \xi_j \rangle_{\mathcal{X}}) = \mathbb{E}(\langle \xi_j, \xi_j \rangle_{\mathcal{X}}).$$

Using this, we have (let now $k > j$)

$$\begin{aligned} \mathbb{E}(\langle \xi_{k+1} - \xi_k, \xi_{j+1} - \xi_j \rangle) &= \mathbb{E}(\langle \xi_{k+1}, \xi_{j+1} \rangle) - \mathbb{E}(\langle \xi_{k+1}, \xi_j \rangle) \\ &\quad - \mathbb{E}(\langle \xi_k, \xi_{j+1} \rangle) + \mathbb{E}(\langle \xi_k, \xi_j \rangle) = 0. \end{aligned}$$

□

Below we sometimes need the assumption that $x \in \mathcal{D}(A)$ almost surely. With Gaussian random variables this means that x is actually a $\mathcal{D}(A)$ -valued random variable.

Proposition 1. *Let ξ be an \mathcal{X} -valued Gaussian random variable s.t. $\xi \in \mathcal{X}_1$ almost surely where $\mathcal{X}_1 \subset \mathcal{X}$ is another Hilbert space with continuous and dense embedding. Then ξ is an \mathcal{X}_1 -valued Gaussian random variable.*

Proof. Pick $h \in \mathcal{X}_1$. We intend to show that $\langle \xi, h \rangle_{\mathcal{X}_1}$ is a real-valued Gaussian random variable. For $h \in \mathcal{X}_1$ there exists $h' \in \mathcal{X}'_1$, the dual space of \mathcal{X}_1 ,

s.t. $\langle \xi, h \rangle_{\mathcal{X}_1} = \langle \xi, h' \rangle_{(\mathcal{X}_1, \mathcal{X}'_1)}$ and further, there exists a sequence $\{h_i\}_{i=1}^\infty \subset \mathcal{X}$ such that $\langle \xi, h' \rangle_{(\mathcal{X}_1, \mathcal{X}'_1)} = \lim_{i \rightarrow \infty} \langle \xi, h_i \rangle_{\mathcal{X}}$. Now $\langle \xi, h_i \rangle_{\mathcal{X}}$ is a pointwise converging sequence of Gaussian random variables and so the limit is also Gaussian. \square

Fernique's theorem [5, Theorem 2.6] can be applied to note that if ξ is an \mathcal{X}_1 -valued Gaussian random variable then $\xi \in L^p(\Omega; \mathcal{X}_1)$ for any $p > 0$. In particular, $\mathbb{E}(\|\xi\|_{\mathcal{X}_1}^2) < \infty$ and if $A \in \mathcal{L}(\mathcal{X}_1, \mathcal{X})$ then $A\xi$ is an \mathcal{X} -valued Gaussian random variable.

2.2. Kalman filter. The discrete time Kalman filter was originally presented in [10]. The continuous time filter is known as the Kalman–Bucy filter, and it was presented in [11]. We also refer to the book [7] for a thorough introduction to both discrete and continuous time Kalman filters as well as the usual techniques needed in different scenarios. Of course, the original presentations are in finite dimensional setting. The infinite dimensional generalization of the discrete time Kalman filter is rather straightforward, and it can be found for example in [9]. The infinite dimensional Kalman–Bucy filter is considered in [3] and [4, Chapter 6]. However, we do not need to be concerned with the continuous time equations. Our approach is based on using the discrete time Kalman filter on a numerable set $\{t_j\}_{j=1}^\infty$ that is dense on an interval $[0, T]$, and bounding the $L^2(\Omega; \mathcal{X})$ -norm of the estimate increment when adding a new time point t_j . In this section we thus review the discrete time Kalman filter equations.

The Kalman filter is based on the fact that with linear systems with Gaussian initial state and input and output noise processes, the state vector remains a Gaussian stochastic process. Also, the conditional expectation of the state with respect to the measurements is a Gaussian process. The statistical properties of the Gaussian \mathcal{X} -valued random variable ξ are completely characterized by the mean $m = \mathbb{E}(\xi) \in \mathcal{X}$ and the covariance operator $P = \text{Cov}[\xi, \xi] \in \mathcal{L}(\mathcal{X})$, defined for $h \in \mathcal{X}$ by $\text{Cov}[\xi, \xi] h := \mathbb{E}((\xi - m) \langle \xi - m, h \rangle_{\mathcal{X}})$. Thus it is meaningful to write $\xi \sim N(m, P)$ meaning that ξ is a Gaussian random variable with mean m and covariance P . The covariance operator is symmetric and nonnegative and, in addition, it is a trace class operator with $\text{tr}(P) = \mathbb{E}(\|\xi - m\|_{\mathcal{X}}^2)$, see [5, Lemma 2.14 & Proposition 2.15]. In fact, by Fernique's theorem, Gaussian random variables are p -integrable for every $p > 0$.

For square integrable random variables, the conditional expectation with respect to a random variable ξ is a projection onto the subspace generated by ξ . With jointly Gaussian random variables $\xi_1 \in \mathcal{X}$ and finite dimensional ξ_2 , this projection has an easy representation. That is, if $\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \sim$

$N\left(\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^* & P_{22} \end{bmatrix}\right)$ then

$$\mathbb{E}(\xi_1 | \xi_2) = m_1 + P_{12} P_{22}^+(\xi_2 - m_2)$$

where P_{22}^+ denotes the (Moore-Penrose) pseudoinverse of P_{22} . The error covariance is

$$\text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_1 - \mathbb{E}(\xi_1|\xi_2)] = P_{11} - P_{12}P_{22}^+P_{12}^*.$$

Now applying the above equations to a Gaussian random variable $[\xi_1, \xi_2, \xi_3]$ where ξ_2 and ξ_3 are finite dimensional, and the 2-by-2 blockwise matrix inversion formula to $\text{Cov} \left[\begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix}, \begin{bmatrix} \xi_2 \\ \xi_3 \end{bmatrix} \right]$ leads directly to

$$(3) \quad \mathbb{E}(\xi_1|[\xi_2, \xi_3]) = \mathbb{E}(\xi_1|\xi_2) + \text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)] \times \\ \times \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)]^+ (\xi_3 - \mathbb{E}(\xi_3|\xi_2))$$

and

$$(4) \quad \text{Cov} [\xi_1 - \mathbb{E}(\xi_1|[\xi_2, \xi_3]), \xi_1 - \mathbb{E}(\xi_1|[\xi_2, \xi_3])] \\ = \text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_1 - \mathbb{E}(\xi_1|\xi_2)] - \text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)] \\ \times \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)]^+ \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_1 - \mathbb{E}(\xi_1|\xi_2)].$$

These equations make it possible to update the state estimate (here $\mathbb{E}(\xi_1|\xi_2)$) recursively when a new measurement (here ξ_3) is obtained from the system.

From (3) we get the covariance for the increment $\mathbb{E}(\xi_1|[\xi_2, \xi_3]) - \mathbb{E}(\xi_1|\xi_2)$,

$$\text{Cov} [\mathbb{E}(\xi_1|[\xi_2, \xi_3]) - \mathbb{E}(\xi_1|\xi_2), \mathbb{E}(\xi_1|[\xi_2, \xi_3]) - \mathbb{E}(\xi_1|\xi_2)] \\ = \text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)] \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)]^+ \\ \times \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_1 - \mathbb{E}(\xi_1|\xi_2)],$$

and further, the $L^2(\Omega; \mathcal{X})$ -norm of the increment is given by

$$(5) \quad \mathbb{E} \left(\|\mathbb{E}(\xi_1|[\xi_2, \xi_3]) - \mathbb{E}(\xi_1|\xi_2)\|_{\mathcal{X}}^2 \right) \\ = \text{tr} \left(\text{Cov} [\xi_1 - \mathbb{E}(\xi_1|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)] \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_3 - \mathbb{E}(\xi_3|\xi_2)]^+ \right. \\ \left. \times \text{Cov} [\xi_3 - \mathbb{E}(\xi_3|\xi_2), \xi_1 - \mathbb{E}(\xi_1|\xi_2)] \right).$$

This fact will be used multiple times in the proofs below.

The familiar discrete time Kalman filter equations follow directly from (3) and (4) if ξ_1 is chosen to be the current state x_i that is to be estimated, ξ_2 consists of the old outputs $[y_1, \dots, y_{i-1}]$, and ξ_3 is the new output y_i .

3. THE CASE WITHOUT INPUT NOISE

For simplicity of presentation, let us first go through the case without input noise. In this case the solution to (1) is simply $z(t) = e^{At}x$.

The convergence rate estimates are based on computing how much $\hat{z}_{T,n}$ can change at most (measured with the $L^2(\Omega; \mathcal{X})$ -norm) when more and more output values $y(t)$ are taken into account from the intervals $t \in ((i-1)T/n, iT/n)$ for $i = 1, \dots, n$. In this section, it is first shown how an

intermediate measurement is taken into account. Consider the output of the system (1), $dy(t) = Ce^{At}x dt + dw(t)$, which is a shortened notation for

$$(6) \quad y(t) = C \int_0^t e^{As} x ds + w(t)$$

where A and C are operators from \mathcal{X} to \mathcal{X} and $\mathcal{Y} = \mathbb{R}^r$, respectively, and w is an r -dimensional Brownian motion with incremental covariance matrix R .

Assume we have a state estimate $\tilde{x}_j := \mathbb{E}(x | \{y(t_1), y(t_2), \dots, y(t_j)\})$ for the initial state x , and the corresponding error covariance $P_j := \text{Cov}[x - \tilde{x}_j, x - \tilde{x}_j]$. Now the next measurement to be taken into account in state estimation is $y(t_{j+1})$. Say $t_{j+1} \in (t_a, t_b)$ for some $a, b \in \{1, \dots, j\}$ and that this interval does not contain any earlier included measurements, that is $t_i \notin (t_a, t_b)$ for $i = 1, \dots, j$. The new state estimate \tilde{x}_{j+1} and the corresponding error covariance $P_{j+1} := \text{Cov}[x - \tilde{x}_{j+1}, x - \tilde{x}_{j+1}]$ are given by (3) and (4), respectively, if we set $\xi_1 = x$, $\xi_2 = [y(t_1), y(t_2), \dots, y(t_j)]$, and $\xi_3 = y(t_{j+1})$.

To get a simple representation for the covariances in (3) and (4), define a new output

$$\tilde{y} := y(t_{j+1}) - \frac{t_b - t_{j+1}}{t_b - t_a} y(t_a) - \frac{t_{j+1} - t_a}{t_b - t_a} y(t_b).$$

That is, \tilde{y} is $y(t_{j+1})$ from which the linear interpolant between $y(t_a)$ and $y(t_b)$ has been removed. By plugging (6) here, this can be written in the form $\tilde{y} = \tilde{C}x + \tilde{w}$ where

$$\begin{aligned} \tilde{C} &= C \int_0^{t_{j+1}} e^{As} ds - C \frac{t_b - t_{j+1}}{t_b - t_a} \int_0^{t_a} e^{As} ds - C \frac{t_{j+1} - t_a}{t_b - t_a} \int_0^{t_b} e^{As} ds \\ &= C \left(\frac{t_b - t_{j+1}}{t_b - t_a} \int_{t_a}^{t_{j+1}} e^{As} ds - \frac{t_{j+1} - t_a}{t_b - t_a} \int_{t_{j+1}}^{t_b} e^{As} ds \right) \end{aligned}$$

and

$$\tilde{w} = w(t_{j+1}) - \frac{t_b - t_{j+1}}{t_b - t_a} w(t_a) - \frac{t_{j+1} - t_a}{t_b - t_a} w(t_b).$$

Since w is Brownian motion, it holds that $\tilde{w} \sim N\left(0, \frac{(t_{j+1} - t_a)(t_b - t_{j+1})}{t_b - t_a} R\right)$ and \tilde{w} is independent of the already included measurements (that is, of ξ_2) and hence of \tilde{x}_j , as well. Thus $\mathbb{E}(\tilde{y} | \xi_2) = \tilde{C}\tilde{x}_j$,

$$\text{Cov}[x - \tilde{x}_j, \tilde{y} - \tilde{C}\tilde{x}_j] = P\tilde{C}^*,$$

and

$$\text{Cov}[\tilde{y} - \tilde{C}\tilde{x}_j, \tilde{y} - \tilde{C}\tilde{x}_j] = \tilde{C}P\tilde{C}^* + \frac{(t_{j+1} - t_a)(t_b - t_{j+1})}{t_b - t_a} R.$$

By (3), the new estimate $\tilde{x}_{j+1} := \mathbb{E}(x | \{y(t_1), y(t_2), \dots, y(t_{j+1})\})$ is given by

$$(7) \quad \tilde{x}_{j+1} = \tilde{x}_j + P_j \tilde{C}^* \left(\tilde{C} P_j \tilde{C}^* + \frac{(t_{j+1} - t_a)(t_b - t_{j+1})}{t_b - t_a} R \right)^{-1} (\tilde{y} - \tilde{C}\tilde{x}_j)$$

and by (4), the new error covariance $P_{j+1} := \text{Cov}[x - \tilde{x}_{j+1}, x - \tilde{x}_{j+1}]$ by

$$(8) \quad P_{j+1} = P_j - P_j \tilde{C}^* \left(\tilde{C} P_j \tilde{C}^* + \frac{(t_{j+1} - t_a)(t_b - t_{j+1})}{t_b - t_a} R \right)^{-1} \tilde{C} P_j.$$

This will be used with $t_b - t_{j+1} = t_{j+1} - t_a = h$, and we define

$$(9) \quad C_h(t)x := \frac{C}{2} \left(\int_{t-h}^t e^{As} x ds - \int_t^{t+h} e^{As} x ds \right), \quad \text{for } t \geq h > 0.$$

Lemma 2. *If $C \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ then for $t \in [h, T-h]$ it holds that*

$$(i) \quad \|C_h(t)\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \leq h\mu \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \quad \text{and}$$

$$(ii) \quad \|C_h(t)\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{Y})} \leq \frac{h^2}{2} \mu \|A\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{X})} \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}.$$

In the finite dimensional case $\|A\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{X})}$ means plainly the matrix norm of A . In the infinite dimensional case $\|A\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{X})} = 1$ because $\mathcal{D}(A)$ is equipped with the graph norm of A .

This could also be shown for more general \tilde{C} with $t_b - t_a$ replacing h in (i) and $\frac{(t_{j+1}-t_a)^2}{2} + \frac{(t_b-t_{j+1})^2}{2}$ replacing h^2 in (ii) but that is not needed. Also, part (ii) can be made a bit better. In fact, $\|C_h(t)x\|_{\mathcal{Y}} \leq \frac{h^2}{2} \mu \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \|Ax\|_{\mathcal{X}}$.

Proof. Part (i) of the Lemma is clear from the definition (9) since $\|e^{At}\|_{\mathcal{L}(\mathcal{X})} \leq \mu$. For part (ii), note that $Ce^{At}x \in C^1(\mathbb{R}^+; \mathcal{Y})$ with $\frac{d}{dt}Ce^{At}x = CAe^{At}x$ and $\|CAe^{At}x\|_{\mathcal{Y}} \leq \mu \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})} \|A\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{X})} \|x\|_{\mathcal{D}(A)}$. Then by Bochner integral properties, C can be taken inside the integral and thus

$$\begin{aligned} & \int_{t-h}^t Ce^{As} x ds - \int_t^{t+h} Ce^{As} x ds \\ &= \int_{t-h}^t \left(Ce^{At}x - \int_s^t CAe^{Ar}x dr \right) ds - \int_t^{t+h} \left(Ce^{At}x + \int_t^s CAe^{Ar}x dr \right) ds \\ &= - \int_{t-h}^t \int_s^t CAe^{Ar}x dr ds - \int_t^{t+h} \int_t^s CAe^{Ar}x dr ds. \end{aligned}$$

This together with the bound for $\|CAe^{At}x\|_{\mathcal{Y}}$ imply (ii). \square

We are now ready to proceed to the actual convergence result which we shall first show in finite dimensional context, namely $\mathcal{X} = \mathbb{R}^p$. The infinite dimensional generalisation will be treated below.

Theorem 1. *Let now $\mathcal{X} = \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times p}$ and $C \in \mathbb{R}^{r \times p}$ (with $r \leq p$) and let $\hat{z}_{T,n}$ and $\hat{z}(T)$ be as defined above in (2), with $u = 0$ in (1). Then*

$$\mathbb{E} \left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2 \right) \leq \frac{MT^3}{n^2}$$

where $M = \frac{\mu^2 \text{tr}(P_0) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2) \|C\|^2 \|A\|^2}{12 \min(\text{eig}(R))}$.

The constant M depends on n through $\mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)$ which is the trace of the error covariance of the discrete time state estimate $\hat{z}_{T,n}$. In order to get

a strict *a priori* result, this term can be bounded by $\mathbb{E}\left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2\right) \leq \mu^2 \text{tr}(P_0)$.

Proof. In the beginning of this section, the output (6) is considered as a signal parameterized by the initial state x and corrupted by noise w . Therefore, it seems beneficial to consider $\mathbb{E}(x|\mathcal{F}_j)$ from which estimates for $z(T)$ can be obtained through $\mathbb{E}(z(T)|\mathcal{F}_j) = e^{AT}\mathbb{E}(x|\mathcal{F}_j)$.

The outline of the proof is as follows. First, we define the martingale $\tilde{z}_j = e^{AT}\tilde{x}_j$ where \tilde{x}_j is the martingale $\tilde{x}_j := \mathbb{E}(x|\mathcal{F}_j)$, where $\mathcal{F}_j = \sigma\{y(t), t \in \mathbb{T}_j\}$ and $\mathbb{T}_j = \{t_i\}_{i=1}^j$ — as explained in Section 2.1. The time points $\{t_i\}_{i=1}^n$ are $t_i = iT/n$ but $\{t_i\}_{i=n+1}^\infty$ will be defined later. The martingales are Gaussian and hence square integrable, and so by Lemma 1, we have the following telescope identity for $L, N \in \mathbb{N}$ with $L \geq N$:

$$(10) \quad \mathbb{E}\left(\|\tilde{z}_L - \tilde{z}_N\|_{\mathcal{X}}^2\right) = \sum_{j=N}^{L-1} \mathbb{E}\left(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2\right).$$

Second, we find an upper bound for $\mathbb{E}\left(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2\right)$ using the results of Section 2.2 and the beginning of this section. Third, we prove that the sum in (10) converges as $L \rightarrow \infty$ and thus \tilde{z}_j is a Cauchy sequence in $L^2(\Omega; \mathcal{X})$. It has a limit in this space by completeness and the limit must be $\hat{z}(T)$ by the considerations in Section 2.1. Also, setting $N = n$ (we have $\tilde{z}_n = \hat{z}_{T,n}$) and letting $L \rightarrow \infty$ in (10) gives $\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right)$.

(I) Martingale \tilde{z}_j : Let $t_i = iT/n$ for $i = 1, \dots, n$. Then \tilde{z}_j for $j = 1, \dots, n$ are the state estimates from the discrete time Kalman filter and, in particular, $\tilde{z}_n = \hat{z}_{T,n}$ defined in (2). The idea is to then halve the intervals $((l-1)T/n, lT/n)$ for $l = 1, \dots, n$ between the already included measurements. That is, we include measurements $y(t_i)$ where $i = n+1, \dots, 2n$, and $t_i = \frac{(i-n-1/2)T}{n}$. Then we halve the new intervals $((l-1)T/2n, lT/2n)$ for $l = 1, \dots, 2n$ by including $2n$ measurements $y(t_i)$ for $i = 2n+1, \dots, 4n$ and $t_i = \frac{(i-2n-1/2)T}{2n}$ and so on. This addition of new time points is illustrated in Fig. 1.

(II) Increment $\tilde{z}_{j+1} - \tilde{z}_j$: Assume that the current state estimate is $\tilde{z}_j = e^{AT}\tilde{x}_j$ with $j \geq n$, the corresponding error covariance matrices are



FIGURE 1. Illustration of the time point addition scheme in the construction of the martingales \tilde{x}_j and \tilde{z}_j .

$e^{AT}P_j e^{A^*T}$ and P_j , respectively, and the next measurement being included is $y(t_{j+1})$ with $j+1 \in \{2^{K-1}n+1, \dots, 2^K n\}$ for some $K = 1, 2, \dots$ (see Fig. 1). Then $t_{j+1} = (2(j - 2^{K-1}n) + 1)h$ with $h = \frac{T}{2^K n}$. The new initial state estimate \tilde{x}_{j+1} is then given by (7) with $\tilde{C} = C_h((2(j - 2^{K-1}n) + 1)h)$ — denoted below simply by C_h — and $h = \frac{T}{2^K n}$. We are only interested in the $L^2(\Omega; \mathcal{X})$ -norm of the \tilde{z} -process increment, and as discussed in Section 2.2, it is obtained from the covariance increment given in (8):

$$\mathbb{E}(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2) = \text{tr} \left(e^{AT} P_j C_h^* (C_h P_j C_h^* + h/2 R)^{-1} C_h P_j e^{A^*T} \right).$$

Now we wish to establish a bound for this trace. To this end, recall that the norm of the inverse of a positive definite matrix is $\|Q^{-1}\| = \frac{1}{\min(\text{eig}(Q))}$, and thus,

$$(11) \quad \left\| \left(C_h P_j C_h^* + \frac{h}{2} R \right)^{-1} \right\| \leq \frac{2}{h \min(\text{eig}(R))} =: \frac{C_R}{h}.$$

Using this and part (ii) of Lemma 2 gives

$$\begin{aligned} & \text{tr} \left(e^{AT} P_j C_h^* \left(C_h P_j C_h^* + \frac{h}{2} R \right)^{-1} C_h P_j e^{A^*T} \right) \\ &= \sum_{k=1}^p \left\langle C_h P_j e^{A^*T} e_k, \left(C_h P_j C_h^* + \frac{h}{2} R \right)^{-1} C_h P_j e^{A^*T} e_k \right\rangle \\ &\leq \frac{C_R}{h} \sum_{k=1}^p \|C_h P_j e^{A^*T} e_k\|_{\mathcal{Y}}^2 = \frac{C_R}{h} \sum_{k=1}^p \|\mathbb{E}(C_h(\tilde{x}_j - x) \langle e^{AT}(\tilde{x}_j - x), e_k \rangle_{\mathcal{X}})\|_{\mathcal{Y}}^2 \\ &\leq \frac{C_R}{h} \mathbb{E}(\|C_h(\tilde{x}_j - x)\|_{\mathcal{Y}}^2) \sum_{k=1}^p \mathbb{E}(\langle e^{AT}(\tilde{x}_j - x), e_k \rangle_{\mathcal{X}}^2) \\ (12) \quad & \leq \frac{C_R}{h} \text{tr}(C_h P_j C_h^*) \mathbb{E}(\|\tilde{z}_j - z(T)\|_{\mathcal{X}}^2) \\ (13) \quad & \leq \frac{h^3}{2 \min(\text{eig}(R))} \mu^2 \|C\|^2 \|A\|^2 \text{tr}(P_j) \mathbb{E}(\|\tilde{z}_j - z(T)\|_{\mathcal{X}}^2). \end{aligned}$$

(III) Convergence: It holds that $\text{tr}(P_j) \leq \text{tr}(P_0)$ and $\mathbb{E}(\|\tilde{z}_j - z(T)\|_{\mathcal{X}}^2) \leq \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)$. In part (II) of the proof we had $h = 2^{-K}T/n$ and that bound is used for all $2^{K-1}n$ new measurements corresponding to this h . Finally, setting $N = n$ and $L \rightarrow \infty$ in (10) and using (13) to bound the terms of the sum yields

$$\begin{aligned} \mathbb{E}(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2) &\leq \sum_{K=1}^{\infty} 2^{K-1}n \left(\frac{T}{2^K n} \right)^3 \frac{\mu^2 \text{tr}(P_0) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2) \|C\|^2 \|A\|^2}{2 \min(\text{eig}(R))} \\ &= \frac{\mu^2 \text{tr}(P_0) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2) \|C\|^2 \|A\|^2 T^3}{12 \min(\text{eig}(R)) n^2} \end{aligned}$$

completing the proof. \square

4. INPUT NOISE

The case with input noise follows exactly the same outline as the case without input noise. The solution to (1) is given by the Wiener integral

$$z(t) = e^{At}x + \int_0^t e^{A(t-s)} B du(s).$$

The idea now is to consider the output $y(t)$ as a process that is parameterized by the initial state x and the input noise process u . To this end, let us define the solution operator $S(t)$ through

$$(14) \quad S(t) : [x, u] \mapsto e^{At}x + \int_0^t e^{A(t-s)} B du(s)$$

and so $z(t) = S(t)[x, u]$. Then the state estimate over a given sigma algebra σ is given by

$$\mathbb{E}(z(T)|\sigma) = S(T)\mathbb{E}([x, u]|\sigma).$$

Hence, we shall virtually consider the estimate of the combined initial state x and the noise process u and then the actual state estimate is obtained through $S(T)$.

Theorem 2. *Let now $\mathcal{X} = \mathbb{R}^p$ and $A \in \mathbb{R}^{p \times p}$ and $C \in \mathbb{R}^{r \times p}$ (with $r \leq p$) and let $\hat{z}_{T,n}$ and $\hat{z}(T)$ be as defined in (2). Then*

$$\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right) \leq \frac{M_1 T^2}{n} + \frac{M_2 T^3}{n^2} + \frac{M_3 T^4}{n^2}$$

where $M_1 = \frac{\|C\|^2 \text{tr}(BQB^*) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)}{\min(\text{eig}(R))}$, $M_2 = \frac{\mu^2 \|A\|^2 \|C\|^2 \text{tr}(P_0) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)}{12 \min(\text{eig}(R))}$,
and $M_3 = \frac{\mu^2 \|C\|^2 \text{tr}(ABQB^*A^*) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)}{2 \min(\text{eig}(R))}$.

As in Theorem 1, an *a priori* result is obtained by bounding $\mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2) \leq \mu^2 \text{tr}(P_0) + T\mu^2 \text{tr}(BQB^*)$.

In the bound of this theorem, the second term originates from the error in the initial state. From the proof below (after (16)) it can be seen that in fact, the different error components can be treated separately (compare (16) with (12)).

Proof. The proof follows exactly the same outline as the proof of Theorem 1. Say we are estimating $[x, u]$ and we have $\tilde{x}_j := \mathbb{E}([x, u]|\mathcal{F}_j)$ and the corresponding error covariance \mathbb{P}_j . Then the state estimate and the corresponding error covariance are given by $\tilde{z}_j = S(T)\tilde{x}_j$ and $S(T)\mathbb{P}_j S(T)^*$ — although in the last equation the formal adjoint $S(T)^*$ is only defined in connection with the covariance \mathbb{P}_j , namely

$$\mathbb{P}_j S(T)^* h = \mathbb{E}((\tilde{x}_j - [x, u]) \langle S(T)(\tilde{x}_j - [x, u]), h \rangle_{\mathcal{X}}).$$

Say we are including measurement $y(t_{j+1})$ which can be written as

$$y(t_{j+1}) = C \int_0^{t_{j+1}} e^{As} x ds + C \int_0^{t_{j+1}} \int_0^s e^{A(s-r)} B du(r) ds + w(t_{j+1}).$$

As in the previous section, to get an output with output noise that is uncorrelated with the already included outputs, we shall subtract the linear interpolant from $y(t_{j+1})$, namely define

$$\tilde{y} = y(t_{j+1}) - \frac{1}{2}y(t_a) - \frac{1}{2}y(t_b)$$

where $t_a = t_{j+1} - h$ and $t_b = t_{j+1} + h$ for proper h . Now this output can be written as

$$\tilde{y} = C_{\square}[x, u]^T + \tilde{w}$$

where $\tilde{w} \sim N(0, h/2R)$ and $C_{\square} = [C_h(t_{j+1}), C_{h,u}(t_{j+1})]$ with $C_h(t_{j+1})$ defined in (9) and

(15)

$$C_{h,u}(t)u := \frac{C}{2} \left(\int_{t-h}^t \int_0^s e^{A(s-r)} B du(r) ds - \int_t^{t+h} \int_0^s e^{A(s-r)} B du(r) ds \right)$$

for $t \geq h$. Now the error covariance increment is as before in (4) and (8) (but with this new output operator C_{\square}) and the $L^2(\Omega; \mathcal{X})$ -norm increment $\mathbb{E}(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2)$ is given by

$$\begin{aligned} \mathbb{E}(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2) &= \text{tr} \left(S(T) \mathbb{P}_j C_{\square}^* \left(C_{\square} \mathbb{P}_j C_{\square}^* + \frac{h}{2} R \right)^{-1} C_{\square} \mathbb{P}_j S(T)^* \right) \\ &= \sum_{k=1}^p \left\langle C_{\square} \mathbb{P}_j S(T)^* e_k, \left(C_{\square} \mathbb{P}_j C_{\square}^* + \frac{h}{2} R \right)^{-1} C_{\square} \mathbb{P}_j S(T)^* e_k \right\rangle_{\mathcal{Y}} \\ &\leq \frac{C_R}{h} \sum_{k=1}^p \|C_{\square} \mathbb{P}_j S(T)^* e_k\|_{\mathcal{Y}}^2 \\ &= \frac{C_R}{h} \sum_{k=1}^p \|\mathbb{E}(C_{\square}(\tilde{x}_j - [x, u]) \langle S(T)(\tilde{x}_j - [x, u]), e_k \rangle_{\mathcal{X}})\|_{\mathcal{Y}}^2 \\ &\leq \frac{C_R}{h} \mathbb{E}(\|C_{\square}(\tilde{x}_j - [x, u])\|_{\mathcal{Y}}^2) \sum_{k=1}^p \mathbb{E}(\langle S(T)(\tilde{x}_j - [x, u]), e_k \rangle_{\mathcal{X}}^2) \\ (16) \quad &\leq \frac{C_R}{h} \mathbb{E}(\|C_{\square}[x, u]\|_{\mathcal{Y}}^2) \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2). \end{aligned}$$

In order to get a suitable bound for the increment, we must find a bound for the term $\mathbb{E}(\|C_{\square}[x, u]\|_{\mathcal{Y}}^2) \leq \mathbb{E}(\|C_h x\|_{\mathcal{Y}}^2) + \mathbb{E}(\|C_{h,u} u\|_{\mathcal{Y}}^2)$ (recall that x and u are independent). As in the proof of Theorem 1, by Lemma 2, the first part is bounded by

$$(17) \quad \mathbb{E}(\|C_h x\|_{\mathcal{Y}}^2) \leq \frac{h^4}{4} \mu^2 \|A\|^2 \|C\|^2 \text{tr}(P_0)$$

so then remains the input noise induced term. To evaluate $C_{h,u}(t_{j+1})u$, note that

$$\int_0^t e^{A(t-s)} B du(s) = \int_0^{t_{j+1}} e^{A(t-s)} B du(s) + \int_{t_{j+1}}^t A \int_0^s e^{A(s-r)} B du(r) ds + \int_{t_{j+1}}^t B du(s)$$

for $t \geq t_{j+1}$ — for $t < t_{j+1}$, just change $t_{j+1} \leftrightarrow t$ in the bounds of the last two integrals and put minus signs in front of them. Of course the last term

is just $\int_{t_{j+1}}^t B du(s) = B(u(t) - u(t_{j+1}))$. Applying this to (15) gives

$$(18) \quad C_{h,u}(t_{j+1})u = -\frac{C}{2} \left[\int_{t_{j+1}-h}^{t_{j+1}} \left(\int_t^{t_{j+1}} A \int_0^s e^{A(s-r)} B du(r) ds + B(u(t) - u(t_{j+1})) \right) dt \right. \\ \left. + \int_{t_{j+1}}^{t_{j+1}+h} \left(\int_{t_{j+1}}^t A \int_0^s e^{A(s-r)} B du(r) ds + B(u(t) - u(t_{j+1})) \right) dt \right].$$

These two terms are very similar by nature so it suffices to find a bound for one of them and use the same bound for both terms. Thus, let us consider the first part of the latter term, namely

$$\begin{aligned} & \frac{C}{2} \int_{t_{j+1}}^{t_{j+1}+h} \int_{t_{j+1}}^t A \int_0^s e^{A(s-r)} B du(r) ds dt \\ &= \frac{C}{2} \int_{t_{j+1}}^{t_{j+1}+h} \int_{t_{j+1}}^t A \int_0^{t_{j+1}} e^{A(s-r)} B du(r) ds dt + \frac{C}{2} \int_{t_{j+1}}^{t_{j+1}+h} \int_{t_{j+1}}^t A \int_{t_{j+1}}^s e^{A(s-r)} B du(r) ds dt \\ &= \frac{C}{2} \int_0^{t_{j+1}} \int_{t_{j+1}}^{t_{j+1}+h} (t_{j+1} + h - s) A e^{A(s-r)} B ds du(r) \\ & \quad + \frac{C}{2} \int_{t_{j+1}}^{t_{j+1}+h} \int_r^{t_{j+1}+h} (t_{j+1} + h - s) A e^{A(s-r)} B ds du(r) \\ &= (I) + (II). \end{aligned}$$

Then

$$\text{Cov}[(I), (I)] = \frac{1}{4} \int_0^{t_{j+1}} \int_{t_{j+1}}^{t_{j+1}+h} \int_{t_{j+1}}^{t_{j+1}+h} (t_{j+1} + h - s)(t_{j+1} + h - r) \\ C e^{A(s-t)} A B Q B^* A^* e^{A^*(r-t)} C^* dr ds dt$$

and from this, using the bound $\|e^{At}\|_{\mathcal{L}(X)} \leq \mu$,

$$\mathbb{E}(\|(I)\|_{\mathbb{Y}}^2) = \text{tr}(\text{Cov}[(I), (I)]) \leq \frac{t_{j+1}h^4}{8} \mu^2 \|C\|^2 \text{tr}(A B Q B^* A^*).$$

For the second term we have

$$\text{Cov}[(II), (II)] = \frac{1}{4} \int_{t_{j+1}}^{t_{j+1}+h} \int_t^{t_{j+1}+h} \int_t^{t_{j+1}+h} (t_{j+1} + h - s)(t_{j+1} + h - r) \\ C A e^{A(t-s)} A B Q B^* A^* e^{A^*(t-r)} C^* dr ds dt$$

and, again,

$$\mathbb{E}(\|(II)\|_{\mathbb{Y}}^2) = \text{tr}(\text{Cov}[(II), (II)]) \leq \frac{h^5}{8} \mu^2 \|C\|^2 \text{tr}(A B Q B^* A^*).$$

In (I), $r \in [0, t_{j+1}]$ and in (II), $r \in [t_{j+1}, t_{j+1} + h]$, and thus they are independent. Using this and the fact that $t_{j+1} + h \leq T$, gives

$$\mathbb{E}(\|(I) + (II)\|_{\mathbb{Y}}^2) \leq \frac{Th^4}{8} \mu^2 \|C\|^2 \text{tr}(A B Q B^* A^*).$$

It is well known that

$$\text{Cov} \left[\int_0^h Bu(t) dt, \int_0^h Bu(t) dt \right] = \frac{h^3}{3} BQB^*$$

and so

$$\mathbb{E} \left(\left\| \frac{C}{2} \int_{t_{j+1}}^{t_{j+1}+h} B(u(t) - u(t_{j+1})) dt \right\|_{\mathcal{Y}}^2 \right) \leq \frac{h^3}{12} \|C\|^2 \text{tr}(BQB^*).$$

In (18), the two $B(u(t) - u(t_{j+1}))$ -terms are independent (because in the first one, $t \leq t_{j+1}$ and in the second, $t \geq t_{j+1}$) and by utilizing this and gathering the above bounds, we get

$$\begin{aligned} \mathbb{E} \left(\|C_{h,u}(t_{j+1})u\|_{\mathcal{Y}}^2 \right) &\leq 6\mathbb{E} \left(\|(I) + (II)\|_{\mathcal{Y}}^2 \right) + 6\frac{h^3}{12} \|C\|^2 \text{tr}(BQB^*) \\ &\leq \frac{3Th^4}{4} \mu^2 \|C\|^2 \text{tr}(ABQB^*A^*) + \frac{h^3}{2} \|C\|^2 \text{tr}(BQB^*). \end{aligned}$$

Combining this with (16) and (17) gives

$$\begin{aligned} \mathbb{E} \left(\|\tilde{z}_{j+1} - \tilde{z}_j\|_{\mathcal{X}}^2 \right) &\leq \frac{h^3}{4} \mu^2 C_R \|A\|^2 \|C\|^2 \text{tr}(P_0) \mathbb{E} \left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2 \right) \\ &\quad + \frac{3Th^3}{4} \mu^2 C_R \|C\|^2 \text{tr}(ABQB^*A^*) \mathbb{E} \left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2 \right) \\ &\quad + \frac{h^2}{2} C_R \|C\|^2 \text{tr}(BQB^*) \mathbb{E} \left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2 \right) \\ &=: C_3 h^3 + C_2 h^2. \end{aligned}$$

Using this bound as in the end of the proof of Theorem 1 gives

$$\mathbb{E} \left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2 \right) \leq \frac{C_2 T^2}{2n} + \frac{C_3 T^3}{6n^2}$$

completing the proof. \square

5. GENERALIZATION TO INFINITE DIMENSIONAL SYSTEMS

We move on to infinite dimensional state space \mathcal{X} . Compared to the finite dimensional case, the main difficulty arises from that the bound for C_h in part (ii) of Lemma 2 utilizes the differentiability of $Ce^{At}x$ and thus it holds for $x \in \mathcal{D}(A)$. A natural assumption that would make it possible to use this bound is that x is a $\mathcal{D}(A)$ -valued random variable. This is exactly what is done in Theorem 4. Before that, in Theorem 3 we shall see, however, that a reasonable convergence estimate can be obtained with slightly less smooth initial state x . Before tackling this problem, we present an example illuminating the necessity of some additional assumptions.

Example 1. This example shows that there is a system with $C \in \mathcal{L}(\mathcal{X}, \mathbb{R})$ such that $\mathbb{E} \left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2 \right)$ converges arbitrarily slowly where $\hat{z}_{T,n}$ and

$\hat{z}(T)$ are defined in (2). Consider the one-dimensional wave equation (without input noise) with augmented state vector,

$$(19) \quad \begin{cases} \frac{d}{dt} \begin{bmatrix} z_1(s, t) \\ z_2(s, t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ \frac{\partial^2}{\partial s^2} & 0 \end{bmatrix} \begin{bmatrix} z_1(s, t) \\ z_2(s, t) \end{bmatrix}, & s \in [0, 1], t \in \mathbb{R}^+, \\ z_1(s, 0) = 0, z_2(s, 0) = x(s), \\ dy(t) = Cz(t) dt + dw(t) \end{cases}$$

in state space $\mathcal{X} = H_0^1[0, 1] \times L^2(0, 1)$ and $\mathcal{D}(A) = (H^2[0, 1] \cap H_0^1[0, 1]) \times H_0^1[0, 1]$. The state is $z(t) = [z_1(t) z_2(t)]^T$. The output operator $C \in \mathcal{L}(\mathcal{X}, \mathbb{R})$ is given by $Cz = \int_0^1 c(s)z_1(s) ds$ where $c(s) = \sum_{k=1}^{\infty} c_k e_k(s)$ with some $\{c_k\} \in l^2$ and $\{e_k\}$ is the orthonormal basis in $L^2(0, 1)$ formed by the sine functions, that is $e_k(s) = \frac{1}{\sqrt{2}} \sin(k\pi s)$. The initial velocity is $x = \sum_{k=1}^{\infty} a_k e_{2k}$ where $a_k \sim N(0, \sigma_k^2)$ and $a_k \perp a_i$ for $k \neq i$. It holds that $\mathbb{E}(\|x\|_{\mathcal{X}}^2) = \sum_{k=1}^{\infty} \sigma_k^2$ and thus this sum is assumed to converge. Then the solution to (19) and the corresponding output are

$$\begin{cases} z_1(s, t) = \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} a_k \sin(2^k \pi s) \sin(2^k \pi t), \\ z_2(s, t) = \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} a_k \sin(2^k \pi s) \cos(2^k \pi t), \\ dy(t) = \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} a_k c_{2^k} \sin(2^k \pi t) dt + dw(t). \end{cases}$$

Now set $T = 1$ and consider the subsequence $\hat{z}_{T, 2^l}$ of the discrete time estimates, defined in (2). As noted in the proof of Thm. 1, it holds that $\mathbb{E}(\|\hat{z}_{T, 2^l} - \hat{z}(T)\|_{\mathcal{X}}^2) = \sum_{i=l}^{\infty} \mathbb{E}(\|\hat{z}_{T, 2^{i+1}} - \hat{z}_{T, 2^i}\|_{\mathcal{X}}^2)$. The estimate $\hat{z}_{T, 2^{l+1}}$ is obtained from the previous estimate $\hat{z}_{T, 2^l}$ by including measurements $y(\frac{2i-1}{2^{l+1}})$ for $i = 1, \dots, 2^l$ as described in the beginning of Section 3. In order to obtain a lower bound for $\mathbb{E}(\|\hat{z}_{T, 2^{l+1}} - \hat{z}_{T, 2^l}\|_{\mathcal{X}}^2)$, define $\hat{C} := [C_h(h), C_h(3h), \dots, C_h(1-h)]^T : \mathcal{X} \rightarrow \mathbb{R}^{2^l}$ where $h = \frac{1}{2^{l+1}}$. That is, \hat{C} gives the whole batch of the measurements needed for the update. For the wave equation it holds that $\|z(t)\|_{\mathcal{X}} = \|x\|_{\mathcal{X}}$ and so the increments $\mathbb{E}(\|\hat{z}_{T, 2^{l+1}} - \hat{z}_{T, 2^l}\|_{\mathcal{X}}^2)$ are the same as the corresponding increments for $\tilde{x}_{2^l} = \mathbb{E}(x | \{y(t), t = j/2^l, j = 1, \dots, 2^l\})$. Then denoting $P_l = \text{Cov}[\tilde{x}_{2^l} - x, \tilde{x}_{2^l} - x]$, it holds that

$$\begin{aligned} \mathbb{E}(\|\hat{z}_{T, 2^{l+1}} - \hat{z}_{T, 2^l}\|_{\mathcal{X}}^2) &= \text{tr} \left(P_l \hat{C}^* \left(\hat{C} P_l \hat{C}^* + \frac{h}{2} R I \right)^{-1} \hat{C} P_l \right) \\ &\geq \left\langle \hat{C} P_l e_{2^{l+1}}, \left(\hat{C} P_l \hat{C}^* + \frac{h}{2} R I \right)^{-1} \hat{C} P_l e_{2^{l+1}} \right\rangle_{\mathbb{R}^{2^l}} \geq \frac{\|\hat{C} P_l e_{2^{l+1}}\|_{\mathbb{R}^{2^l}}^2}{\max(\text{eig}(\hat{C} P_l \hat{C}^* + \frac{h}{2} R I))}. \end{aligned}$$

For $h = 2^{-l}$ it holds that $C_h(ih)e_{2^k} = 0$ when $l < k$ and $i = 1, \dots, 2^l - 1$ because when computing $C_h(ih)e_{2^k}$ by (9), the integrals are always over full periods of the sine function $\sin(2^k \pi t)$. When $l = k$ it holds that $C_h(ih)e_{2^k} = \frac{\sqrt{2}h}{\pi} c_{2^k}$ for every $i = 1, 3, \dots, 2^k - 1$. So, loosely speaking, the already included output values $y(\frac{2i-1}{2^l})$ do not carry any information on a_k for $k > l$.

Thus $P_l e_{2^{l+1}} = \sigma_{l+1}^2 e_{2^{l+1}}$ and $\|\widehat{C} P_l e_{2^{l+1}}\|_{\mathbb{R}^{2^l}}^2 = 2^l \sigma_{l+1}^2 \left(\frac{\sqrt{2}h}{\pi} c_{2^{l+1}} \right)^2$. For the denominator it holds by part (i) of Lemma 2 that

$$\max \left(\text{eig} \left(\widehat{C} P_l \widehat{C}^* + \frac{h}{2} R I \right) \right) \leq \frac{h}{2} R + \mathbb{E} \left(\|\widehat{C} x\|_{\mathbb{R}^{2^l}}^2 \right) \leq \frac{h}{2} R + 2^l h^2 \|C\|_{\mathcal{L}(\mathcal{X}, \mathbb{R})}^2 \text{tr}(P_0).$$

Recalling $h = \frac{1}{2^{l+1}}$, we finally get $\mathbb{E}(\|\hat{z}_{T, 2^{l+1}} - \hat{z}_{T, 2^l}\|_{\mathcal{X}}^2) \geq \frac{8\sigma_{l+1}^2 c_{2^{l+1}}^2}{\pi^2 R + 4\pi^2 \|C\|_{\mathcal{L}(\mathcal{X}, \mathbb{R})}^2 \text{tr}(P_0)}$ and further

$$\mathbb{E}(\|\hat{z}_{T, 2^l} - \hat{z}(T)\|_{\mathcal{X}}^2) \geq \frac{8 \sum_{i=l+1}^{\infty} \sigma_i^2 c_{2^i}^2}{\pi^2 R + 4\pi^2 \|C\|_{\mathcal{L}(\mathcal{X}, \mathbb{R})}^2 \text{tr}(P_0)}$$

where there is no h -dependence and the variances $\{\sigma_k^2\}$ can be chosen so that the convergence is arbitrarily slow, concluding the example.

Clearly some additional assumptions are needed for getting any convergence rate estimates. In the following theorem, the initial state is assumed to be so smooth that the covariance operator satisfies $P_0 \in \mathcal{L}(\mathcal{X}, \mathcal{D}(A))$. As noted after Theorem 1, the error components stemming from the initial state and the input noise can be treated separately. Therefore, the following two theorems treat the noiseless case and the input noise is treated in Corollary 1.

Theorem 3. *Let $\hat{z}_{T,n}$ and $\hat{z}(T)$ be as defined in (2) with $u = 0$ in (1), and assume $C \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Assume $x \sim N(m, P_0)$ where the covariance operator satisfies $P_0 \in \mathcal{L}(\mathcal{X}, \mathcal{D}(A))$. Then*

$$\mathbb{E}(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2) \leq \frac{MT^2}{n}$$

where $M = \frac{r \|P_0\|_{\mathcal{L}(\mathcal{X}, \mathcal{D}(A))} \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}^2 \mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2)}{2 \min(\text{eig}(R))}$. Recall that

$$\mathbb{E}(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2) \leq \mu^2 \text{tr}(P_0).$$

Proof. The main idea of the proof is the same as in the proof of Theorem 1 and we note that every step taken until equation (12) in that proof can be taken in the infinite dimensional setting as well — p just has to be replaced by ∞ in the sums but this does not cause any problems.

So we pick up from (12) and note first that

$$\begin{aligned} \text{tr}(C_h P_j C_h^*) &\leq r \|C_h P_j C_h^*\|_{\mathcal{L}(\mathcal{Y})} = r \sup_{\|y\|_{\mathcal{Y}}=1} \langle y, C_h P_j C_h^* y \rangle_{\mathcal{Y}} \\ &= r \sup_{\|y\|_{\mathcal{Y}}=1} \langle C_h^* y, P_j C_h^* y \rangle_{\mathcal{X}} \leq r \sup_{\|y\|_{\mathcal{Y}}=1} \langle C_h^* y, P_0 C_h^* y \rangle_{\mathcal{X}} = r \|C_h P_0 C_h^*\|_{\mathcal{L}(\mathcal{Y})} \end{aligned}$$

where $r = \dim(\mathcal{Y})$. The inequality $P_j \leq P_0$ was used in \mathcal{X} , but now the $\mathcal{L}(\mathcal{X}, \mathcal{D}(A))$ -norm can be used for P_0 . Then using both parts (i) and (ii) of Lemma 2 gives

$$\|C_h P_0 C_h^*\|_{\mathcal{L}(\mathcal{Y})} \leq \frac{h^3}{2} \mu^2 \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}^2 \|P_0\|_{\mathcal{L}(\mathcal{X}, \mathcal{D}(A))}.$$

As before, this leads to an estimate

$$\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right) \leq \frac{r\mu^2 \|P_0\|_{\mathcal{L}(\mathcal{X}, \mathcal{D}(A))} \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}^2 \mathbb{E}\left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2\right) T^2}{2 \min(\text{eig}(R)) n} =: \frac{MT^2}{n}$$

completing the proof. \square

Checking the assumption $P_0 \in \mathcal{L}(\mathcal{X}, \mathcal{D}(A))$ might be difficult. Under the stronger smoothness assumption $x \in \mathcal{D}(A)$ almost surely, we get the same convergence rate as in the finite dimensional case:

Theorem 4. *Make the same assumptions as in Theorem 3. Assume, in addition, that $x \in \mathcal{D}(A)$ almost surely. Then*

$$\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right) \leq \frac{MT^3}{n^2}$$

$$\text{where } M = \frac{\mu^2 \text{tr}(AP_0A^*) \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}^2 \mathbb{E}\left(\|\hat{z}_{T,n} - z(T)\|_{\mathcal{X}}^2\right)}{12 \min(\text{eig}(R))}.$$

Proof. The proof is the same as that of Theorem 1 but from Eq. (12) we proceed differently. It holds that

$$\text{tr}(C_h P_j C_h^*) \leq \text{tr}(C_h P_0 C_h^*) = \mathbb{E}\left(\|C_h x\|_{\mathcal{Y}}^2\right) \leq \frac{h^4}{4} \mu^2 \|C\|_{\mathcal{L}(\mathcal{X}, \mathcal{Y})}^2 \mathbb{E}\left(\|Ax\|_{\mathcal{X}}^2\right)$$

where the last inequality holds by part (ii) of Lemma 2. The term is finite by Proposition 1 and Fernique's theorem. Further, it holds that $\mathbb{E}\left(\|Ax\|_{\mathcal{X}}^2\right) = \text{tr}(AP_0A^*)$. Now the result follows as above. \square

As discussed after Theorem 2, the error components stemming from the initial state error and the input noise can be treated separately. Therefore, as an almost direct corollary of Theorems 2, 3, and 4, we obtain the following result:

Corollary 1. *Let $\hat{z}_{T,n}$ and $\hat{z}(T)$ be as defined in (2) and assume $C \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ and $B \in \mathcal{L}(\mathcal{U}, \mathcal{D}(A))$. Assume also either (i): $P_0 \in \mathcal{L}(\mathcal{X}, \mathcal{D}(A))$, or (ii): $x \in \mathcal{D}(A)$ almost surely. Then*

$$\mathbb{E}\left(\|\hat{z}_{T,n} - \hat{z}(T)\|_{\mathcal{X}}^2\right) \leq \frac{M_1 T^2}{n} + \frac{M_3 T^4}{n^2} + \text{err}_x$$

where M_1 and M_3 are as in Theorem 2 and err_x is as in Theorem 3 in the case of assumption (i), or as in Theorem 4 in the case of assumption (ii).

The proof is the same as the proof of Theorem 2, with the modifications of Theorems 3 or 4. Note that $\text{tr}(ABQB^*A^*) \leq \|A\|_{\mathcal{L}(\mathcal{D}(A), \mathcal{X})}^2 \|B\|_{\mathcal{L}(\mathcal{U}, \mathcal{D}(A))}^2 \text{tr}(Q)$ and $\text{tr}(BQB^*) \leq \|B\|_{\mathcal{L}(\mathcal{U}, \mathcal{X})}^2 \text{tr}(Q)$.

6. DISCUSSION

Since the implementation of the discrete time Kalman filter is straightforward, it is a tempting choice for state estimation for discretized continuous time systems. As the temporal discretization is refined, the discrete time state estimate converges pointwise to the continuous time estimate in $L^2(\Omega; \mathcal{X})$. In this article, we derived convergence speed estimates at which

the discrete time Kalman filter estimate converges to the continuous time estimate as the temporal discretization is refined. The result was achieved for both finite and infinite dimensional systems with bounded observation operator and smooth input operator. In the case of infinite dimensional systems, some smoothness assumption on the initial state is needed for obtaining any convergence speed estimates. This was demonstrated in Example 1. Possible additional assumptions are (i): for the initial state covariance it holds that $P_0 \in \mathcal{L}(\mathcal{X}, \mathcal{D}(A))$; or (ii): for the initial state it holds that $x \in \mathcal{D}(A)$ almost surely. In the latter case we obtained the same convergence speed estimate as for finite dimensional systems.

A topic that would require further work are systems with infinite dimensional output space. The output space dimension r does not appear explicitly in the convergence speed estimates, except for Thm. 3. However, in the proofs we need an upper bound for $\left\| (C_h P_j C_h^* + \frac{h}{2} R)^{-1} \right\|_{\mathcal{L}(\mathcal{Y})}$ and thus, in order to obtain (11), we made a coercivity assumption $R \geq \epsilon I > 0$ which excludes infinite dimensional output space since R is required to be a trace class operator. In the beginning, we also assumed that the input space \mathcal{U} is finite dimensional. This is merely an assumption by which tedious definitions of infinite dimensional Wiener processes are avoided. For more on this subject, we refer to [5].

Two more topics that are not covered by this article are the long time behaviour as $T \rightarrow \infty$, and using some approximate time integration scheme for taking the time step. When T grows, the error covariance converges under some assumptions on the observability of the system. Of course, the observability of the continuous time system does not imply the observability of the discretized system. In the case where there is input noise affecting the system, the error covariance limits are obtained as the solutions P_d and P_c of the corresponding discrete or continuous time algebraic Riccati equations, respectively. Then it holds that $\lim_{n \rightarrow \infty} \mathbb{E}(\|\hat{x}_{n\Delta t, n} - \hat{x}(n\Delta t)\|_{\mathcal{X}}^2) = \text{tr}(P_d - P_c)$ where $\hat{x}_{n\Delta t, n}$ and $\hat{x}(n\Delta t)$ are defined in (2). Finally, further research would be needed to study the error caused to the state estimate if some numerical time integration scheme is used for computing the discrete time update, that is, $e^{A\Delta t}$ is not computed accurately. A similar problem is addressed in [2] and [18], but they are mainly concerned with the stability of the resulting filter.

Acknowledgements. The author was financially supported by the Finnish Graduate School in Engineering Mechanics. The author thanks Dr. Jarmo Malinen for valuable comments on the manuscript.

REFERENCES

- [1] A. Aalto, Spatial discretization error in Kalman filtering for discrete-time infinite dimensional systems, ArXiv:1406.7160 (2014), 19 pages.
- [2] P. Axelsson and F. Gustafsson, Discrete-time solutions to the continuous-time differential Lyapunov equation with applications to Kalman filtering, IEEE T. Automat. Contr., 60 (2015), 632–643.
- [3] A. Bensoussan, “Filtrage Optimal des Systèmes Linéaires,” Dunod, Paris, 1971.
- [4] R. Curtain and A. Pritchard, “Infinite Dimensional Linear Systems Theory,” Springer-Verlag, 1979.

- [5] G. Da Prato and J. Zabczyk, “Stochastic Equations in Infinite Dimensions,” Encyclopedia of Mathematics and its Applications 44, Cambridge University Press, 1992.
- [6] M. H. A. Davis, “Linear Estimation and Stochastic Control,” Wiley, New York, 1977.
- [7] A. Gelb, “Applied Optimal Estimation,” MIT Press, Cambridge, MA, 1974.
- [8] A. Germani, L. Jetto, and M. Piccioni, Galerkin approximation for optimal linear filtering of infinite-dimensional linear systems, *SIAM J. Control Optim.*, 26 (1988), 1287–1305.
- [9] L. L. Horowitz, “Optimal Filtering of Gyroscopic Noise,” PhD. thesis, Massachusetts Institute of Technology, 1974.
- [10] R. Kalman, A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82 (1960), 35–45.
- [11] R. Kalman and R. Bucy, New results in linear filtering and prediction theory, *Journal of Basic Engineering*, 83 (1961), 95–107.
- [12] W. Lee, D. McDougall, and A. Stuart, Kalman filtering and smoothing for linear wave equations with model error, *Inverse Problems*, 27 (2011).
- [13] B. O. Øksendal, “Stochastic Differential Equations: An Introduction with Applications,” Springer-Verlag, 1998.
- [14] G. Pisier, “Martingales in Banach Spaces (in Connection with Type and Cotype),” Manuscript, Course IHP, Feb. 2–8, 2011. Available at: <http://www.math.jussieu.fr/~pisier/ihp-pisier.pdf>
- [15] M. Salgado, R. Middleton, and G. Goodwin, Connection between continuous and discrete Riccati equations with applications to Kalman filtering, *IEE Proceedings*, 135 (1988), 28–34.
- [16] D. Simon, “Optimal State Estimation — Kalman, H_∞ , and Nonlinear Approaches,” John Wiley & Sons, 2006.
- [17] J. Sun, Sensitivity analysis of the discrete-time algebraic Riccati equation, *Linear Algebra Appl.*, 275–276 (1998), 595–615.
- [18] N. Wahlström, P. Axelsson, and F. Gustafsson, Discretizing stochastic dynamical systems using Lyapunov equations, *arXiv:1402.1358* (2014), 17 pages.